Available online at www.sciencedirect.com
ScienceDirect
Jo nal of M l i a a e Anal i 98 (2007) 40 56
Journal of Multivariate Analysis
.el e ie com/loca e/jm a

A mp o ic di ib ion of nonpa ame ic eg e ion e ima o fo longi dinal o f nc ional da a

Fang Yao
Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA

Recei ed 28 Jan a 2005
A ailable online 28 Sep embe 2006

Abstract

The e ima ion of a eg e ion f nc ion b ke el me hod fo longi dinal o f nc ional da a i con ide ed. In he con e of longi dinal da a anal i , a andom f nc ion picall ep e en a bjec ha i of en ob e ed a a mall n mbe of ime poin , hile in he die of f nc ional da a he andom eali a ion i all mea ed on a den e g id. Ho e e , e en iall he ame me hod can be applied o bo h ampling plan , a ell a in a n mbe of e ing l ing be een hem. In hi pape gene al e l a e de i ed fo he a mp o ic di ib ion of eal- al ed f nc ion i h a g men hich a e f nc ional fo med b eigh ed a e age of longi dinal o f nc ional da a. A mp o ic di ib ion fo he e ima o of he mean and co a iance f nc ion ob ained f om noi ob e a ion i h he p e ence of i hin- bjec co ela ion a e died. The e a mp o ic no mali e l a e compa able o ho e anda d a e ob ained f om independen da a, hich i ill a ed in a im la ion d . Be ide , hi pape di c e he condi ion a ocia ed i h ampling plan , hich a e eq i ed fo he alidi of local p ope ie of ke el-ba ed e ima o fo longi dinal o f nc ional da a.

AMS 1991 subject classification: 62G08

Keywords: A mp o ic di ib ion; Co a iance; F nc ional da a; Longi dinal da a; Reg e ion; Wi hin- bjec co ela ion

1. Introduction

Mode echnolog and ad anced comp ing en i onmen ha e facili a ed he collec ion and anal i of high-dimen ional da a, o da a ha a e epea edl mea ed fo a ample of bjec . The epea ed mea emen a e of en eco ded o e a pe iod of ime, a on an clo ed and bo nded in e al $\mathcal{T}$. I al o co ld be a pacial a iable, ch a in image o geo cience applica ion .

E-mail addresses: f ao@ a . o on o.ed , f ao@ a .colo a e.ed .

0047-259X/$ - ee f on ma e 2006 El e ie Inc. All igh e e ed.
doi:10.1016/j.jm a.2006.08.007

When the data are recorded densely over time, often by machine, they are typically termed functional or curve data with one observed curve or function per subject, while in longitudinal studies the repeated measurements usually take place on a few scattered observational time points for each subject. A significant intrinsic difference between two settings lies in the perception that functional data are observed in the continuum without noise [2,3], whereas longitudinal data are observed at sparsely distributed time points and are often subject to experimental error [4]. However, in practice functional data are analyzed after smoothing noisy observations [10], which indicates that the difference between two data types related to the way in which a problem is perceived are arguably more conceptual than actual. Therefore in this paper, kernel-based regression estimators obtained from observations at discrete time points contaminated with measurement errors, rather than observations in the continuum, are considered for the realistic reason. In the context of kernel-based nonparametric regression, the effects of sampling plans on the statistical estimators are also investigated.

A vast literature has been developed in the past decade on the kernel-based regression for independent and identically distributed data, for summary, see Fan and Gijbels [5]. There has been substantial recent interest in extending the existing asymptotic results to functional or longitudinal data [8,11,14,13,9]. The issues caused by the within-subject correlation are rigorously addressed in this paper. Hart and Wehrly [8] studied the Gasser–Müller estimator of the mean function for repeated measurements observed on a regular grid by assuming stationary correlation structure, and showed that the influence of the within-subject correlation on the asymptotic variance is of smaller order compared to the standard rate obtained from independent data and will disappear when the correlation function is differentiable at zero. Our asymptotic distribution result is in fact consistent with that in Hart and Wehrly [8] and applicable for general covariance structure without stationary assumption. This problem was also discussed by Staniswalis and Lee [12] and Lin and Carroll [9], where they used the heuristic arguments of the local property of local polynomial estimation and in this way ignored the within-subject correlation while deriving the asymptotic variances. This paper derives appropriate conditions that are required for the validity of the local property of kernel type estimators obtained from longitudinal or functional data. These conditions also provide practical guidelines for various sampling procedures.

The contribution of this paper is the derivation of general asymptotic distribution results in both one-dimensional and two-dimensional smoothing contexts for real-valued functions with arguments which are functionals formed by weighted averages of longitudinal or functional data. These asymptotic normality results are comparable to those obtained for identically distributed and independent data. These results are applied to the kernel-based estimators of the mean and covariance functions, which yield asymptotic normal distributions of these estimators. In particular, to the best of our knowledge, no asymptotic distribution results are available up to date for nonparametric estimation of covariance functions obtained from longitudinal or functional data contaminated with measurement errors. By comparison, Hall et al. [6,7] investigated asymptotic properties of nonparametric kernel estimators of autocovariance, where the measurements were only observed from a single stationary stochastic process or random field. Although the asymptotic distributions are derived for random designs in this paper, the arguments can be extended to fixed designs and other sampling plans with appropriate modifications, and asymptotic bias and variance terms can also be obtained in similar manner. This will provide theoretical basis and practical guidance for the nonparametric analysis of functional or longitudinal data with important potential applications which are based on the asymptotic distributions. Typical examples include the construction of asymptotic confidence bands for regression functions and confidence regions for co ariance rface , and al o fa elec ion of band id h for co ariance rface e ima ion ba ed on a mp o ic mean q ared error. O her applica ion in he con e of moo hing indepen-den da a can be e plored for he moo hing of longi dinal or f nc ional da a ing kernel-ba ed e ima or.

The remainder of he paper i organi ed a follo . In Sec ion 2 e deri e he general a mp o ic di rib ion of one- and o-dimen ional moo her ob ained from longi dinal or f nc ional da a for random de ign. The e general a mp o ic re l are applied o commonl ed kernel- pe e ima or of he mean c r e and co ariance rface in Sec ion 3. E en ion o fi ed de ign i di c ed in Sec ion 4. A im la ion d i pre en ed o e al a e he deri ed a mp o ic re l for correla ed da a in Sec ion 5, hile di c ion , incl ding po en ial applica ion of he re l ing a mp o ic normali , are offered in Sec ion 6.

2. General results of asymptotic distributions for random design

In hi ec ion e ill define general f nc ional ha are kernel- eigh ed a erage of he da a for one-dimen ional and o-dimen ional moo hing. The in rod ced general f nc ional incl de he mo commonl ed pe of kernel-ba ed e ima or a pecial ca e , ch a Ga er–M ller e ima or, Nadara a–Wa on e ima or, local pol nomial e ima or, e c. Since Nadara a–Wa on and local pol nomial e ima or are mo l ed in prac ice, heir a mp o ic beha ior in erm of bia and ariance for independen da a ha e been horo ghl died in e i ing li era re. Ho e er, for longi dinal or f nc ional da a, par ic larl in regard o co ariance rface e ima or, he a mp o ic beha ior of bia and ariance of he e o commonl ed e ima or are ill largel nkno n. Therefore in Sec ion 3, he general a mp o ic re l de eloped in hi ec ion are applied o Nadara a–Wa on and local pol nomial e ima or in bo h one-dimen ional and o-dimen ional moo hing e ing . In par ic lar, he lack of a mp o ic re l for he co ariance rface e ima or of longi di-nal or f nc ional da a i an addi ional mo i a ion for he defini ion of he o-dimen ional general f nc ional ha can be applied o de elop he a mp o ic di rib ion for he e e ima or.

We fir con ider random de ign hile e en ion o o her ampling plan i deferred o Sec ion 4. In cla ical longi dinal die , mea remen are of en in ended o be on a reg lar ime grid. Ho e er, ince indi id al ma mi ched led i i , he re l ing da a all become par e, here onl fe ob er a ion are ob ained for mo bjec , i h neq al n mber of repea ed mea remen per bjec and differen mea remen ime T_{ij} per indi id al. Thi ampling variance σ^2,

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(T_{ij}) + \varepsilon_{ij}, \quad T_{ij} \in \mathcal{T}, \tag{1}$$

where $E\varepsilon_{ij} = 0$, $var(\varepsilon_{ij}) = \sigma^2$, and the number of observations $N_i(n)$ depending on the sample size n, are considered random. We make the following assumptions,

(A1.1) The number of observations $N_i(n)$ made for the ith subject or cluster, $i = 1, \ldots, n$, is a r.v. with $N_i(n) \overset{\text{i.i.d}}{\sim} N(n)$, where $N(n) > 0$ is a positive integer valued random variable with $\limsup_{n\to\infty} EN(n)^2/[EN(n)]^2 < \infty$ and $\limsup_{n\to\infty} EN(n)^4/EN(n)EN(n)^3 < \infty$.

In the sequel the dependence of $N_i(n)$ and $N(n)$ on the sample size n is suppressed for implicity; i.e., $N_i = N_i(n)$ and $N(n) = N$. The observation times and measurements are assumed to be independent of the number of measurements, i.e., for any subset $J_i \subseteq \{1, \ldots, N_i\}$ and for all $i = 1, \ldots, n$,

(A1.2) $(\{T_{ij} : j \in J_i\}, \{Y_{ij} : j \in J_i\})$ is independent of N_i.

Writing $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{iN_i})^T$ and $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$, it is easy to see that the triples $\{\boldsymbol{T}_i, \boldsymbol{Y}_i, N_i\}$ are i.i.d..

2.1. Asymptotic normality of one-dimensional smoother

To assume appropriate regularity conditions that are used to derive asymptotic properties, we define a new type of continuity that differs from those which are commonly used. We say that a real function $f(x, y) : \Re^{p+q} \to \Re$ is continuous on $x \in A \subseteq \Re^p$ uniformly in $y \in \Re^q$, provided that for any $x \in A$ and $\varepsilon > 0$, there exists a neighborhood of x not depending on y, saying $U(x) \subseteq \Re^p$, such that $|f(x', y) - f(x, y)| < \varepsilon$ for all $x' \in U(x)$ and $y \in \Re^q$.

For random design, (T_{ij}, Y_{ij}) are assumed to have the identical distribution as (T, Y) with joint density $g(t, y)$. Assume that the observation times T_{ij} are i.i.d. with the marginal density $f(t)$, but dependence is allowed among Y_{ij} and Y_{ik} that are observations made for the same subject or cluster. Also denote the joint density of (T_j, T_k, Y_j, Y_k) by $g_2(t_1, t_2, y_1, y_2)$, where $j \neq k$. Let v, k be given integers, with $0 \leqslant v < k$. We assume regularity conditions for the marginal and joint densities, $f(t)$, $g(t, y)$, $g_2(t_1, t_2, y_1, y_2)$ and the mean function of the underlying process $X(t)$, i.e., $E[X(t)] = \mu(t)$, with respect to a neighborhood of an interior point $t \in \mathcal{T}$, assuming that there exists a neighborhood $U(t)$ of t such that:

(B1.1) $\frac{d^k}{du^k} f(u)$ exists and is continuous on $u \in U(t)$, and $f(u) > 0$ for $u \in U(t)$;

(B1.2) $g(u, y)$ is continuous on $u \in U(t)$ uniformly in $y \in \Re$; $\frac{d^k}{du^k} g(u, y)$ exists and is continuous on $u \in U(t)$ uniformly in $y \in \Re$;

(B1.3) $g_2(u, v, y_1, y_2)$ is continuous on $(u, v) \in U(t)^2$ uniformly in $(y_1, y_2) \in \Re^2$;

(B1.4) $\frac{d^k}{du^k} \mu(u)$ exists and is continuous on $u \in U(t)$.

Let $K_1(\cdot)$ be nonnegative univariate kernel functions in one-dimensional smoothing. The assumptions for kernel $K_1 : \Re \to \Re$ are as follows. We say that a univariate kernel function K_1 is of order (v, k), if

$$\int u^\ell K_1(u)\, du = \begin{cases} 0, & 0 \leqslant \ell < k, \quad \ell \neq v, \\ (-1)^v v!, & \ell = v, \\ \neq 0, & \ell = k, \end{cases} \tag{2}$$

(B2.1) K_1 is compactly supported, $\|K_1\|^2 = \int K_1^2(u)\,du < \infty$;
(B2.2) K_1 is a kernel function of order (v, ℓ).

Let $b = b(n)$ be a sequence of bandwidths that are used in one-dimensional smoothing. We develop asymptotics as $n \to \infty$, and require
(B3) $b \to 0$, $n(EN)b^{v+1} \to \infty$, $b(EN) \to 0$, and $n(EN)b^{2k+1} \to d^2$ for some d with $0 \leqslant d < \infty$.
One could see in the proof of Theorem 1 that the assumptions (B3) combined with (A1.1) provide the conditions such that the local properties of kernel-type estimators hold for longitudinal or functional data with the presence of within-subject correlation.

Let $\{\psi_\lambda\}_{\lambda=1,\ldots,l}$ be a collection of real functions $\psi_\lambda : \Re^2 \to \Re$, which satisfy:

(B4.1) $\psi_\lambda(t, y)$ are continuous on $\{t\}$ uniformly in $y \in \Re$;
(B4.2) $\frac{d^k}{dt^k}\psi_\lambda(t, y)$ exist for all arguments (t, y) and are continuous on $\{t\}$ uniformly in $y \in \Re$.

Then we define the general weighted averages

$$\Psi_{\lambda n} = \frac{1}{nENb^{v+1}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \psi_\lambda(T_{ij}, Y_{ij}) K_1\left(\frac{t - T_{ij}}{b}\right), \quad \lambda = 1, \ldots, l.$$

and

$$\mu_\lambda = \mu_\lambda(t) = \frac{d^v}{dt^v} \int \psi_\lambda(t, y) g(t, y)\, dy, \quad \lambda = 1, \ldots, l.$$

Let

$$\sigma_{\kappa\lambda} = \sigma_{\kappa\lambda}(t) = \int \psi_\kappa(t, y)\psi_\lambda(t, y) g(t, y)\, dy \|K_1\|^2, \quad 1 \leqslant \lambda, \kappa \leqslant l,$$

and $H : \Re^l \to \Re$ be a function with continuous first order derivatives. We denote the gradient vector $((\partial H/\partial x_1)(v), \ldots, (\partial H/\partial x_l)(v))^T$ by $DH(v)$ and $\bar{N} = \sum_{i=1}^n N_i/n$.

Theorem 1. *If the assumptions* (A1.1), (A1.2) *and* (B1.1)–(B4.2) *hold, then*

$$\sqrt{n\bar{N}b^{2v+1}}[H(\Psi_{1n}, \ldots, \Psi_{ln}) - H(\mu_1, \ldots, \mu_l)] \xrightarrow{\mathcal{D}} \mathcal{N}(\beta, [DH(\mu_1, \ldots, \mu_l)]^T \Sigma[DH(\mu_1, \ldots, \mu_l)]), \tag{3}$$

where

$$\beta = \frac{(-1)^k d}{k!} \int u^k K_1(u)\, du \sum_{\lambda=1}^{l} \frac{\partial H}{\partial \mu_\lambda}\{(\mu_1, \ldots, \mu_l)^T\} \frac{d^{k-v}}{dt^{k-v}} \mu_\lambda(t), \quad \Sigma = (\sigma_{\kappa\lambda})_{1 \leqslant \kappa, \lambda \leqslant l}.$$

Proof. It is seen that $\bar{N}$ can be replaced with EN by Slutsky Theorem under (A1.1). We now show that

$$\sqrt{n(EN)b^{2v+1}}[H(E\Psi_{1n}, \ldots, E\Psi_{ln}) - H(\mu_1, \ldots, \mu_l)] \longrightarrow \beta. \tag{4}$$

Since (A1.1) and (A1.2) hold, and K_1 i of order (ν, k), ing Ta lore pan ion o order k, one ob ain

$$
\begin{aligned}
E\Psi_{\lambda n} &= \frac{1}{nb^{\nu+1}} E\left\{\sum_{i=1}^{n} \frac{1}{EN} \sum_{j=1}^{N_i} \psi_\lambda(T_{ij}, Y_{ij}) K_1\left(\frac{t-T_{ij}}{b}\right)\right\} \\
&= \frac{1}{b^{\nu+1}EN} E\left\{\sum_{j=1}^{N} E\left[\psi_\lambda(T_j, Y_j) K_1\left(\frac{t-T_j}{b}\right)\Bigg| N\right]\right\} \\
&= \frac{1}{b^{\nu+1}} E\left\{\psi_\lambda(T, Y) K_1\left(\frac{t-T}{b}\right)\right\} \\
&= \mu_\lambda + \frac{(-1)^k}{k!} \int u^k K_1(u)\, du \, \frac{d^{k-\nu}}{dt^{k-\nu}} \mu_\lambda(t) b^{k-\nu} + o(b^{k-\nu}).
\end{aligned} \tag{5}
$$

Then (4) follo from an l-dimen ional Ta lore pan ion of H of order 1 aro nd $(\mu_1, \ldots, \mu_l)^T$, co pled i h (5). If e can ho

$$
\sqrt{n(EN)b^{2\nu+1}}[(\Psi_{1n}, \ldots, \Psi_{ln})^T - (E\Psi, \ldots, E\Psi_{ln})^T] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma), \tag{6}
$$

in analog o Bha achara and M ller [1], and con in i of DH a $(\mu_1, \ldots, \mu_l)^T$ and appl - ing imilar arg men ed in (5), e find $DH(E\Psi_{1n}, \ldots, E\Psi_{ln}) \to DH(\mu_1, \ldots, \mu_l)$. Then Carmèr-Wold de ice ield

$$
\sqrt{n(EN)b^{2\nu+1}}[H(\Psi_{1n}, \ldots, \Psi_{ln}) - H(E\Psi, \ldots, E\Psi_{ln})] \xrightarrow{\mathcal{D}} \mathcal{N}(0, DH(\mu_1, \ldots, \mu_l)^T \Sigma DH(\mu_1, \ldots, \mu_l)), \tag{7}
$$

combined i h (4), leading o (3).

I remain o ho (6). Ob er ing (A1.1) and (A1.2), one ha

$$
\begin{aligned}
& n(EN)b^{2\nu+1} cov(\Psi_{\lambda n}, \Psi_{\kappa n}) \\
&= \frac{1}{b} E\left\{\frac{1}{EN}\left[\sum_{j=1}^{N} \psi_\lambda(T_j, Y_j) K_1\left(\frac{t-T_j}{b}\right)\right]\left[\sum_{k=1}^{N} \psi_\kappa(T_k, Y_k) K_1\left(\frac{t-T_k}{b}\right)\right]\right\} \\
&\quad - \frac{EN}{b} E\left[\frac{1}{EN} \sum_{j=1}^{N} \psi_\lambda(T_j, Y_j) K_1\left(\frac{t-T_j}{b}\right)\right] \\
&\quad \times E\left[\frac{1}{EN} \sum_{k=1}^{N} \psi_\kappa(T_k, Y_k) K_1\left(\frac{t-T_k}{b}\right)\right] \\
&\equiv I_1 - I_2.
\end{aligned}
$$

I i ob io ha $I_2 = O(b) = o(1)$ f om he de a ion of (5). Fo I_1, i can be en a

$$I_1 = \frac{1}{b} E\left[\frac{1}{EN}\sum_{j=1}^{N} \psi_\lambda(T_j, Y_j)\psi_\kappa(T_j, Y_j)K_1^2\left(\frac{t - T_j}{b}\right)\right]$$

$$+\frac{1}{b} E\left[\frac{1}{EN}\sum_{1\leqslant j\neq k\leqslant N} \psi_\lambda(T_j, Y_j)\psi_\kappa(T_k, Y_k)K_1\left(\frac{t - T_j}{b}\right)K_1\left(\frac{t - Y_k}{b}\right)\right]$$

$$\equiv Q_1 + Q_2.$$

Appl ing (A1.1) and (A1.2), one ha

$$Q_1 = \frac{1}{b} E\left\{\frac{1}{EN}\sum_{j=1}^{N} E\left[\psi_\lambda(T_j, Y_j)\psi_\kappa(T_j, Y_j)K_1^2\left(\frac{t - T_j}{b}\right)\middle| N\right]\right\}$$

$$= \frac{1}{b} E\left[\psi_\lambda(T, Y)\psi_\kappa(T, Y)K_1^2\left(\frac{t - Y}{b}\right)\right] = \sigma_{\lambda\kappa} + o(1).$$

Then (4) ill hold, ob e ing (A1.1) and he follo ing a g men ha g a an ee he local p ope of he ke nel-ba ed e ima o i h he p e ence of i hin- bjec co ela ion in longi dinal o f nc ional da a,

$$Q_2 = \frac{1}{bEN} E\left\{\sum_{1\leqslant j\neq k\leqslant N}^{N} E\left[\psi_\lambda(T_j, Y_j)\psi_\kappa(T_k, Y_k)K_1\left(\frac{t - T_j}{b}\right)K_1\left(\frac{t - T_k}{b}\right)\middle| N\right]\right\}$$

$$= \frac{EN(N-1)}{bEN} E\left[\psi_\lambda(T_1, Y_1)\psi_\kappa(T_2, Y_2)K_1\left(\frac{t - T_1}{b}\right)\right]K_1\left(\frac{t - T_2}{b}\right)$$

$$= \frac{bEN(N-1)}{EN}\int_{\Re^4} \psi_\lambda(t - ub, y_1)\psi_\kappa(t - vb, y_2)K_1(u)K_2(v)$$

$$\times g_2(t - ub, t - vb, y_1, y_2)\,du\,dv\,dy_1\,dy_2$$

$$= \frac{bEN(N-1)}{EN}\int_{\Re^2} \psi_\lambda(t, y_1)\psi_\kappa(t, y_2)g_2(t, t, y_1, y_2)\,dy_1\,dy_2 + o(b) = o(1),$$

i.e., he i hin- bjec co ela ion can be igno ed hile de ing he a mp o ic a iance. □

2.2. *Asymptotic normality of two-dimensional smoother*

The gene al a mp o ic e l can be e ended o o-dimen ional moo hing. Le $(\boldsymbol{v}, \boldsymbol{k})$ deno e he m l i-indice $\boldsymbol{v} = (v_1, v_2)$ and $\boldsymbol{k} = (k_1, k_2)$, he e $|\boldsymbol{v}| = v_1 + v_2$ and $|\boldsymbol{k}| = k_1 + k_2$. In o-dimen ional moo hing, mo e eg la a mp ion a e needed fo join den i ie . Le $f_2(s, t)$ be he join den i of (T_j, T_k), and $g_4(s, t, s', t', y_1, y_2, y_1', y_2')$ he join den i of $(T_j, T_k, T_{j'}, T_{k'}, Y_j, Y_k, Y_{j'}, Y_{k'})$ he e $j \neq k$, $(j, k) \neq (j', k')$. Deno e he co a iance face b $C(s, t) = cov(X(T_j), X(T_k)|T_j = s, T_k = t)$. The follo ing eg la condi ion a e a med, he e $U(s, t)$ i ome neighbo hood of $\{(s, t)\}$,

(C1.1) $\frac{d^{|\boldsymbol{k}|}}{du^{k_1}\,dv^{k_2}} f_2(u, v)$ e i and i con in o on $(u, v) \in U(s, t)$, and $f_2(u, v) > 0$ fo $(u, v) \in U(s, t)$;

(C1.2) $g_2(u, v, y_1, y_2)$ is continuous on $(u, v) \in U(s, t)$ uniformly in $(y_1, y_2) \in \Re^2$; $\frac{d^{|\boldsymbol{k}|}}{du^{k_1}\, dv^{k_2}} g_2(u, v, y_1, y_2)$ exists and is continuous on $(u, v) \in U(s, t)$ uniformly in $(y_1, y_2) \in \Re^2$;

(C1.3) $g_4(u, v, u', v', y_1, y_2, y_1', y_2')$ is continuous on $(u, v, u', v') \in U(s, t)^2$ uniformly in $(y_1, y_2, y_1', y_2') \in \Re^4$;

(C1.4) $\frac{d^{|\boldsymbol{k}|}}{du^{k_1}\, dv^{k_2}} C(u, v)$ exists and is continuous on $(u, v) \in U(s, t)$.

Let K_2 be nonnegative bivariate kernel function used in the two-dimensional smoothing. The assumptions for kernel K_2 are as follows,

(C2.1) K_2 is compact supported with $\|K_2\|^2 = \int_{\Re^2} K_2^2(u, v)\, du\, dv < \infty$, and is symmetric with respect to coordinate u and v.

(C2.2) K_2 is a kernel function of order $(|\boldsymbol{v}|, |\boldsymbol{k}|)$, i.e.,

$$\sum_{\ell_1+\ell_2=|\boldsymbol{l}|} \int_{\Re^2} u^{\ell_1} v^{\ell_2} K_2(u, v)\, du\, dv = \begin{cases} 0, & 0 \leqslant |\boldsymbol{l}| < |\boldsymbol{k}|, |\boldsymbol{l}| \neq |\boldsymbol{v}|, \\ (-1)^{|\boldsymbol{v}|} |\boldsymbol{v}|!, & |\boldsymbol{l}| = |\boldsymbol{v}|, \\ \neq 0, & |\boldsymbol{l}| = |\boldsymbol{k}|. \end{cases} \tag{8}$$

Let $h = h(n)$ be a sequence of bandwidth used in two-dimensional smoothing, while it is possible that the bandwidths used for two arguments may be different. Since we will focus on the estimator of the covariance surface that is symmetric about the diagonal, it is sufficient to consider the identical bandwidth for the two arguments. The asymptotic is developed as $n \to \infty$ as follows:

(C3) $h \to 0$, $nEN^2 h^{|\boldsymbol{v}|+2} \to \infty$, $hEN^3 \to 0$, and $nE[N(N-1)]h^{2|\boldsymbol{k}|+2} \to e^2$ for some $0 \leqslant e < \infty$.

Similar to the one-dimensional smoothing case, assumptions (C3) and (A1.1) guarantee the local properties of the bivariate kernel-based estimator with the presence of within-subject correlation.

Let $\{\phi_\lambda\}_{\lambda=1,\ldots,l}$ be a collection of real functions $\phi_\lambda : \Re^4 \to \Re$, $\lambda = 1, \ldots, l$, satisfying

(C4.1) $\phi_\lambda(s, t, y_1, y_2)$ are continuous on $\{(s, t)\}$ uniformly in $(y_1, y_2) \in \Re^2$;

(C4.2) $\frac{d^{|\boldsymbol{k}|}}{ds^{k_1} dt^{k_2}} \phi_\lambda(s, t, y_1, y_2)$ exist for all arguments (s, t, y_1, y_2) and are continuous on $\{(s, t)\}$ uniformly in $(y_1, y_2) \in \Re^2$.

Then the general weighted averages of two-dimensional smoothing are defined by, for $1 \leqslant \lambda \leqslant l$,

$$\Phi_{\lambda n} = \Phi_{\lambda n}(t, s) = \frac{1}{nE[N(N-1)]h^{|\boldsymbol{v}|+2}} \sum_{i=1}^{n} \sum_{1 \leqslant j \neq k \leqslant N_i} \phi_\lambda(T_{ij}, T_{ik}, Y_{ij}, Y_{ik}) \times K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right).$$

Let

$$m_\lambda = m_\lambda(s, t) = \sum_{v_1+v_2=|\boldsymbol{v}|} \frac{d^{|\boldsymbol{v}|}}{ds^{v_1}\, dt^{v_2}} \int_{\Re^2} \phi_\lambda(s, t, y_1, y_2) g_2(s, t, y_1, y_2)\, dy_1\, dy_2, \quad 1 \leqslant \lambda \leqslant l,$$

and

$$\omega_{\kappa\lambda} = \omega_{\kappa\lambda}(s,t) = \int_{\Re^2} \phi_\kappa(s,t,y_1,y_2)\phi_\lambda(s,t,y_1,y_2) g_2(s,t,y_1,y_2) dy_1 dy_2 \|K_2\|^2,$$

$$1 \leqslant \kappa, \lambda \leqslant l,$$

and $H : \Re^l \to \Re$ is a function with continuous first-order derivatives as previously defined.

Theorem 2. *If assumptions* (A1.1), (A1.2) *and* (C1.1)–(C4.2) *hold*, *then*

$$\sqrt{n\bar{N}(\bar{N}-1)h^{2|\mathbf{v}|+2}}[H(\Phi_{1n},\ldots,\Phi_{ln}) - H(m_1,\ldots,m_l)]$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(\gamma, [DH(m_1,\ldots,m_l)]^T \Omega [DH(m_1,\ldots,m_l)]), \tag{9}$$

where

$$\gamma = \frac{(-1)^{|\boldsymbol{k}|} e}{|\boldsymbol{k}|!} \sum_{\lambda=1}^{l} \left\{ \sum_{k_1+k_2=|\boldsymbol{k}|} \int_{\Re^2} u^{k_1} v^{k_2} K_2(u,v)\, du\, dv \frac{d^{|\boldsymbol{k}|}}{ds^{k_1}\, dt^{k_2}} \right.$$

$$\left. \times \int_{\Re^2} \phi_\lambda(s,t,y_1,y_2) g_2(s,t,y_1,y_2)\, dy_1\, dy_2 \right\}$$

$$\times \left\{ \frac{\partial H}{\partial m_\lambda}(m_1,\ldots,m_l)^T \right\},$$

$$\Omega = (\omega_{\kappa\lambda})_{1 \leqslant \kappa \leqslant l}.$$

The proof of Theorem 2 essentially follows that of Theorem 1 with appropriate modifications which are required for two-dimensional smoothing.

3. Applications to nonparametric regression estimators for functional or longitudinal data

Although various versions of kernel-based estimators have been introduced in literature, Nadaraya–Watson and local polynomial, especially local linear estimators, are the most commonly used nonparametric smoothing techniques in longitudinal or functional data analysis. Due to within-subject correlation, the asymptotic behavior in terms of bias and variance of these estimators for noisily observed longitudinal or functional data have been as well understood as for i.i.d. data. Especially, asymptotic results for covariance estimators do not exist. Therefore in this section, we apply the asymptotic results developed for general functional to Nadaraya–Watson and local linear estimators of regression function and covariance surface to obtain their asymptotic distributions.

3.1. Asymptotic distributions of mean estimators

We apply Theorem 1 to the local asymptotic distribution of the commonly used Nadaraya–Watson kernel estimator $\hat{\mu}_N(t)$ and local linear estimator $\hat{\mu}_L(t)$ for functional/longitudinal da a:

$$\hat{\mu}_{\mathrm{N}}(t) = \left[\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_1\left(\frac{t-T_{ij}}{b}\right) Y_{ij}\right] \Bigg/ \left[\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_1\left(\frac{t-T_{ij}}{b}\right)\right], \tag{10}$$

$$\hat{\mu}_{\mathrm{L}}(t) = \hat{\alpha}_0(t) = \arg\min_{(\alpha_0,\,\alpha_1)} \left\{\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_1\left(\frac{t-T_{ij}}{b}\right)[Y_{ij} - (\alpha_0 + \alpha_1(T_{ij} - t))]^2\right\}. \tag{11}$$

Corollary 1. *If assumptions* (A1.1), (A1.2), *and* (B1.1) (B3) *hold with* $v = 0$ *and* $k = 2$, *then*

$$\sqrt{n\bar{N}b}[\hat{\mu}_{\mathrm{N}}(t) - \mu(t)] \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{d}{2}\frac{\mu^{(2)}(t)f(t) + 2\mu^{(1)}(t)f^{(1)}(t)}{f(t)}\sigma_{K_1}^2, \frac{var(Y|T=t)\|K_1\|^2}{f(t)}\right), \tag{12}$$

where d is as in (B3), $\sigma_{K_1}^2 = \int u^2 K_1(u)\,du$ *370TD(2/F5295Tj/-0.9121TD/F11Tf5.9776005.9776 TJ.1T3F*

Here $w_{ij} = K_1((t - T_{ij})/b)/(nb)$, here K_1 is a kernel function of order $(0, 2)$, satisfying (B2.1) and (B2.2), and $\hat{\alpha}_1(t)$ is an estimator for the first derivative $\mu'(t)$ of μ at t.

Observing that Corollary 1 implies $\hat{\mu}_{\rm N}(t) \xrightarrow{p} \mu(t)$, let $\hat{f}(t) = \sum_i \sum_j w_{ij}/N_i$, it is easy to show $\hat{f}(t) \xrightarrow{p} f(t)$ in analogy to Corollary 1. We proceed to show $\hat{a}_1(t) \xrightarrow{p} \mu'(t)$. Denote $\sigma^2_{K_1} = \int u^2 K_1(u)\, du$, the kernel function $\widetilde{K}_1(t) = -tK_1(t)/\sigma^2_{K_1}$, and define $\Psi_{\lambda n}$, $1 \leqslant \lambda \leqslant 3$ by $\psi_1(u, y) = y$, $\psi_2(u, y) \equiv 1$, $\psi_3(u, y) = u - t$. Observe that $\widetilde{K}_1$ is of order $(1, 3)$, $\hat{f}(t) \xrightarrow{p} f(t)$, and define

$$\widetilde{H}(x_1, x_2, x_3) = \frac{x_1 - x_2\hat{\mu}_{\rm N}(t)}{x_3 - bx_2^2/\hat{f}(t)\cdot\sigma^2_{K_1}} \quad \text{and} \quad H(x_1, x_2, x_3) = \frac{x_1 - x_2\mu(t)}{x_3}.$$

Then

$$\hat{\alpha}_1(t) = \widetilde{H}(\Psi_{1n}, \Psi_{2n}, \Psi_{3n}) = \left[H(\Psi_{1n}, \Psi_{2n}, \Psi_{3n}) + \frac{\Psi_{2n}(\mu(t) - \hat{\mu}_{\rm N}(t))}{\Psi_{3n}}\right]\frac{\Psi_{3n}}{\Psi_{3n} + b^2\Psi_{2n}^2/\hat{f}(t)\cdot\sigma^2_{K_1}}.$$

Note that $\mu_1 = (\mu' f + mf')(t)$, $\mu_2 = f'(t)$, and $\mu_3 = f(t)$, implying $\Psi_{\lambda n} - \mu_\lambda = O_p(1/\sqrt{n\bar{N}b^3})$, for $\lambda = 1, 2, 3$, by Theorem 1. Using *Slutsky's* Theorem, $|\widetilde{H}(\Psi_{1n}, \Psi_{2n}, \Psi_{3n}) - \mu'(t)| = O_p(1/\sqrt{n\bar{N}b^3})$ follows.

For the asymptotic distribution of $\hat{\mu}_{\rm L}$, note that

$$\hat{\mu}_{\rm L}(t) = \frac{\sum_i \frac{1}{EN}\sum_j w_{ij}Y_{ij} - \sum_i \frac{1}{EN}\sum_j w_{ij}(T_{ij} - t)\hat{a}_1(t)}{\sum_i \frac{1}{EN}\sum_j w_{ij}}.$$

Considering $\sqrt{n\bar{N}b}\sum_i \frac{1}{EN}\sum_j w_{ij}(T_{ij} - t) = \sqrt{n\bar{N}b}\sigma^2_{K_1}b^2\Psi_{2n}$. Since $\tilde{K}_1$ is of order $(1, 3)$, Theorem 1 implies $\Psi_{2n} = f'(t) + O_p(1/\sqrt{n\bar{N}b^3})$, which yields $\sqrt{n\bar{N}b}\sigma^2_{K_1}b^2\Psi_{2n} = \sqrt{n\bar{N}b^5}\sigma^2_{K_1} f'(t) + \sigma^2_{K_1}O_p(b) = o_p(1)$ by observing $n\bar{N}b^5 \to d^2$ for $0 \leqslant d < \infty$. Since $\hat{f}(t) \xrightarrow{p} f(t)$ and $|\hat{\alpha}_1(t) - \mu'(t)| = O_p(1/\sqrt{n\bar{N}b^3}) = o_p(1)$, we find

$$\lim_{n\to\infty}\sqrt{n\bar{N}b}[\hat{\mu}_{\rm L}(t) - \mu(t)] \stackrel{\mathcal{D}}{=} \lim_{n\to\infty}\sqrt{n\bar{N}b} \times \left\{\frac{\sum_i \frac{1}{EN}\sum_j w_{ij}Y_{ij} - \mu'(t)\sum_i \frac{1}{EN}\sum_j w_{ij}T_{ij} + t\mu'(t)\sum_i \frac{1}{EN}\sum_j w_{ij}}{\sum_i \frac{1}{EN}\sum_j w_{ij}} - \mu(t)\right\}.$$

Using the kernel K_1 of order $(0, 2)$, we re-define $\Psi_{\lambda n}$, $1 \leqslant \lambda \leqslant 3$, through $\psi_1(u, y) = y$, $\psi_2(u, y) = u$ and $\psi_3(u, y) \equiv 1$, setting $v = 0$, $k = 2$, $l = 3$ and $H(x_1, x_2, x_3) = [x_1 - \mu'(t)x_2 + t\mu'(t)x_3]/x_3$. Then (13) follows by applying Theorem 1. $\square$

3.2. Asymptotic distributions of covariance estimators

Note that in model (1), $cov(Y_{ij}, Y_{ik}|T_{ij}, T_{ik}) = cov(X(T_{ij}), X(T_{ik})) + \sigma^2\delta_{jk}$, where δ_{jl} is 1 if $j = k$ and 0 otherwise. Let $C_{ijk} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{ik} - \hat{\mu}(T_{ik}))$ be the raw covariances, where $\hat{\mu}(t)$ is the estimated mean function obtained from the previous step, for instance, $\hat{\mu}(t) = \hat{\mu}_{\rm N}(t)$ or $\hat{\mu}(t) = \hat{\mu}_{\rm L}(t)$. It is easy to see that $E[C_{ijk}|T_{ij}, T_{ik}] \approx cov(X(T_{ij}), X(T_{ik})) + \sigma^2\delta_{jk}$. Therefore, the diagonal of the raw covariance should be removed, i.e., only C_{ijk}, $j \neq k$, should be included as input data for the covariance surface smoothing step, as previously observed in Staniswalis and Lee [12] and Yao et al. [15].

Commonly used nonparametric regression estimators of the covariance surface, $C(s,t) = E\{[X(T_1) - \mu(T_1)][X(T_2) - \mu(T_2)|T_1 = s, T_2 = t]\}$, are the two-dimensional Nadaraya Watson estimator and local linear estimator defined as follows:

$$\widehat{C}_{\mathrm{N}}(s,t) = \left[\sum_{i=1}^{n} \sum_{j \neq k} K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right) C_{ijk}\right] \Bigg/ \left[\sum_{i=1}^{n} \sum_{j \neq k} K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right)\right], \tag{16}$$

$$\widehat{C}_{\mathrm{L}}(s,t) = \hat{\beta}_0(s,t) = \arg\min_{\boldsymbol{\beta}} \left\{\sum_{i=1}^{n} \sum_{j \neq k} K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right)\right.$$

$$\times [C_{ijk} - f(\boldsymbol{\beta}, (s,t), (T_{ij}, T_{ik}))]$$

$\phi_1(t_1, t_2, y_1, y_2) = (y_1 - \mu(t_1))(y_2 - \mu(t_2))$, $\phi_2(t_1, t_2, y_1, y_2) = y_1 - \mu(t_1)$, and $\phi_3(t_1, t_2, y_1, y_2) \equiv 1$, hen p$_{t,s\in\mathcal{T}} |\Phi_{pn}| = O_p(1)$, fo $p = 1, 2, 3$, b Lemma 1 of Yao e al. [16]. Thi implie ha p$_{t,s\in\mathcal{T}} |\Phi_{2n}| O_p(1/(\sqrt{n}b)) = O_p(1/(\sqrt{n}b))$ and p$_{t,s\in\mathcal{T}} |\Phi_{3n}| O_p(1/(\sqrt{n}b)) = O_p(1/(\sqrt{n}b))$. Since p$_{t\in\mathcal{T}} |\hat{\mu}(t) - \mu(t)|^2 = O_p(1/(nb))$ a e negligible compa ed o Φ_{1n}, he Nada a Wa on e ima o $\widehat{C}_{\mathrm{N}}(s,t)$, of $C(s,t)$ ob ained f om C_{ijk} i a mp o icall eq i alen o ha ob ained f om $\tilde{C}_{ijk}$, deno ed b $\tilde{C}_{\mathrm{N}}(t,s)$.

The efo e, i i fficien o ho ha he a mp o ic di ib ion of $\tilde{C}_{\mathrm{N}}(s,t)$ follo (18). Choo e $\boldsymbol{v} = (0,0)$, $|\boldsymbol{k}| = 2$, $\phi_1(s,t,y_1,y_2) = (y_1 - \mu(s))(y_2 - \mu(t))$, $\phi_2(s,t,y_1,y_2) \equiv 1$ and $H(x_1,x_2) = x_1/x_2$ in Theo em 2, hen $\tilde{C}_{\mathrm{N}}(s,t) = H(\Psi_{1n}, \Psi_{2n})$. To comp e $\gamma_{\mathrm{N}}(s,t)$, e $DH(m_1,m_2) = (1/m_2, -m_1/m_2^2)$, and no e $m_1(s,t) = \int_{\Re^2}(y_1-\mu(s))(y_2-\mu(t))g_2(s,t,y_1,y_2)\,dy_1dy_2 = f_2(s,t)C(s,t)$ and $m_2(s,t) = f_2(s,t)$. One ha $(d^2/dt^2)m_1(s,t) = [(d^2f_2/dt^2)C + 2(df_2/dt)(dC/dt) + f_2(d^2C/dt^2)](s,t)$, $(d^2/d^2t)m_2(s,t) = d^2f_2(s,t)/dt^2$ and imila de a- i e i h e pec o he a g men s leading o he bia e m in (12). Fo he a mp o ic a ance, no e ha $\omega_{11} = \|K_2\|^2 \int_{\Re^2}(y_1-\mu(s))^2(y_2-\mu(t))^2 g_2(s,t,y_1,y_2)dy_1dy_2 = E[(Y_1 - \mu(T_1))^2(Y_2 - \mu(T_2))^2 | T_1 = s, T_2 = t) f_2(s,t)\|K_2\|^2$, $\omega_{12} = \omega_{21} = \|K_2\|^2 f_2(s,t)C(s,t)$, $\omega_{22} = \|K_2\|^2 f_2(s,t)$, and $DH(m_1,m_2) = (1/m_2, -m_1/m_2^2)$, ielding he a ance e m in (12). □

Corollary 4. *If the assumptions* (A1.1), (A1.2), *and* (C1.1) (C3) *hold with* $|\boldsymbol{v}| = 0$ *and* $|\boldsymbol{k}| = 2$, *then*

$$\sqrt{n\bar{N}(\bar{N}-1)h^2}[\widehat{C}_{\mathrm{L}}(s,t) - C(s,t)]$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{e}{4}\sigma_{K_2}^2[d^2C(s,t)/ds^2 + d^2C(s,t)/dt^2], \frac{v(s,t)\|K_2\|^2}{f_2(s,t)}\right), \tag{19}$$

where e is as in (C3), $v(s,t) = var\{(Y_1 - \mu(T_1))(Y_2 - \mu(T_2))|T_1 = s, T_2 = t)$, $\sigma_{K_2}^2 = \int_{\Re^2}(u^2 + v^2)K_2(u,v)\,du\,dv$, $\|K_2\|^2 = \int_{\mathcal{R}^2} K_2^2(u,v)\,du\,dv$.

Proof. In analog o he p oof of Co olla 3, he local linea e ima o $\widehat{C}_{\mathrm{L}}(s,t)$ ob ained f om C_{ijk} i a mp o icall eq i alen o ha ob ained f om $\tilde{C}_{ijk}$, deno ed b $\tilde{C}_{\mathrm{L}}(t,s)$. Al o deno e he ol ion o (17), af e b i ing $\tilde{C}_{ijk}$ fo C_{ijk}, b $\tilde{\boldsymbol{\beta}}(s,t) = (\tilde{\beta}_0(s,t), \tilde{\beta}_1(s,t), \tilde{\beta}_2(s,t))$, and in fac $\tilde{\beta}_0(s,t) = \tilde{C}_{\mathrm{L}}(s,t)$. Fo implici , le $W_{ijk} = K_2((s - T_{ij})/h, (t - T_{ik})/h)/(nh^2)$ and "$\sum_{i,j\neq k}$" i abb e ia ion of "$\sum_{i=1}^n \sum_{j\neq k}$". Algeb a calc la ion ield ha

$$\tilde{C}_{\mathrm{L}} = \frac{\sum_{i,j\neq k}\tilde{C}_{ijk}W_{ijk} - \tilde{\beta}_1\sum_{i,j\neq k}W_{ijk}T_{ij} + \tilde{\beta}_1\sum_{i,j\neq k}W_{ijk}s - \tilde{\beta}_2\sum_{i,j\neq k}W_{ijk}T_{ik} + \tilde{\beta}_2\sum_{i,j\neq k}W_{ijk}t}{\sum_{i,j\neq k}W_{ijk}},$$

$$\tilde{\beta}_1 = \frac{R_{00}(S_{10}S_{02} - S_{01}S_{11}) + R_{10}(S_{00}S_{02} - S_{01}S_{20}) - R_{01}(S_{00}S_{11} - S_{10}S_{02})}{S_{00}S_{20}S_{02} - S_{00}S_{11}^2 - S_{10}^2S_{02} + S_{10}S_{01}S_{11} + S_{20}S_{10}S_{11} - S_{01}S_{20}^2},$$

$$\tilde{\beta}_2 = \frac{R_{00}(S_{10}S_{11} - S_{01}S_{02}) - R_{10}(S_{00}S_{11} - S_{01}S_{20}) + R_{01}(S_{00}S_{20} - S_{10}^2)}{S_{00}S_{20}S_{02} - S_{00}S_{11}^2 - S_{10}^2S_{02} + S_{10}S_{01}S_{11} + S_{20}S_{10}S_{11} - S_{01}S_{20}^2},$$

he e

$$R_{pq} = \sum_{i,j\neq k} W_{ijk}(T_{ij} - s)^p(T_{ik} - t)^q\tilde{C}_{ijk}, \quad S_{pq} = \sum_{i,j\neq k} W_{ijk}(T_{ij} - s)^p(T_{ik} - t)^q.$$

No e ha $\tilde{\beta}_1$ and $\tilde{\beta}_2$ a e local linea e ima o of he pa ial de a i e of $C(s,t)$, $dC(s,t)/ds$ and $dC(s,t)/dt$, e pec i el . In analog o he p oof of Co olla 2, i can be ho n ha $|\tilde{\beta}_1(s,t) - dC(s,t)/ds| = O_p(1/\sqrt{nEN(N-1)h^4})$ and $|\tilde{\beta}_2(s,t) - dC(s,t)/dt| = O_p(1/\sqrt{n\bar{N}(\bar{N}-1)h^4})$ b appl ing Theo em 2. Then one can b i e $dc(s,t)/ds$, $dC(s,t)/dt$ fo $\tilde{\beta}_1(s,t)$, $\tilde{\beta}_2(s,t)$ in $\tilde{C}_{\mathrm{L}}(s,t)$, and deno e he l ing e ima o b $C_{\mathrm{L}}^*(s,t)$. I i ea o ee ha

$$\lim_{n\to\infty} \sqrt{n\bar{N}(\bar{N}-1)h^2}[C_{\mathrm{L}}(s,t) - C(s,t)] \stackrel{\mathcal{D}}{=} \lim_{n\to\infty} \sqrt{n\bar{N}(\bar{N}-1)h^2}[C_{\mathrm{L}}^*(s,t) - C(s,t)].$$

We define $\Phi_{\lambda n}$, $1 \leqslant \lambda \leqslant 4$, h gh $\phi_1(s,t,y_1,y_2) = (y_1 - \mu(s))(y_2 - \mu(t))$, $\phi_2(s,t,y_1,y_2)$hand.1(h gh)]TJ2 those in Corollaries 3 and 4, with $f(t)$ replaced by $1/|\mathcal{T}|$ and $f(s,t)$ replaced by $1/|\mathcal{T}|^2$, where $|\mathcal{T}|$ is the length of the interval.

5. Simulation study

A numerical study is conducted to evaluate the derived asymptotic properties. The key finding in this paper is that the asymptotic results for functional/longitudinal are comparable to those obtained from independent data, i.e., the influence of within-subject covariance does not play significant role in determining the asymptotic bias and variance. For simplicity, we focus on the local polynomial mean estimators which are often superior to the Nadaraya–Watson estimators.

We first generated $M = 200$ samples consisting of $n = 50$ i.i.d. random trajectories each. Following model (1), the simulated process has a mean function $\mu(t) = (t-1/2)^2$, $0 \leqslant t \leqslant 1$ which has a constant second derivative $\mu^{(2)}(t) = 2$, and a constant within-subject covariance function derived from a random intercept $\xi_1 \overset{\text{i.i.d.}}{\sim} N(0, \lambda_1)$, where $\lambda_1 = 0.01$ and $\phi_1(t) = 1$, $0 \leqslant t \leqslant 1$. The measurement errors in (1) are assumed to be $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, where $\sigma^2 = 0.01$. A random design was used, where the numbers of observations for each subject N_i were chosen from $\{2, 3, 4, 5\}$ with equal likelihood and the locations of the observations were uniformly distributed on $[0, 1]$, i.e., $T_{ij} \overset{\text{i.i.d.}}{\sim} U[0, 1]$. For comparison, we generated $M = 200$ samples of $n = 50$ i.i.d. random trajectories which have the same structures as in model (1) but no within-subject correlation. Letting $\xi_{i1} = 0$ and $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sqrt{\lambda_1 + \sigma^2})$ leads to independent data with the same mean and variance functions. Therefore, the two sets of data have the same asymptotic distribution for the local polynomial mean estimators. We also generated $M = 200$ correlated and independent samples, respectively, consisting of $n = 200$ trajectories each for demonstrating the asymptotic behavior with the increasing sample size n.

Here we use the Epanechnikov kernel function, i.e., $K_1(u) = 3/4(1-u^2)\mathbf{1}_{[-1,1]}(u)$, where $\mathbf{1}_A(u) = 1$ if $u \in A$ and 0 otherwise for any set A. Note that $n(EN)b^{2k+1} \to d^2$ in (B3), $\mu^{(2)}(t) = 2$, $var(Y|T = t) = \lambda_1 + \sigma^2 = 0.02$, and the design density $f(t) = 1$, where $k = 2$ for local polynomial estimators and b is the bandwidth used for the mean estimation. From the above construction, one can calculate the asymptotic variance and bias of the local polynomial mean estimator $\mu_{\mathrm{L}}(t)$ using Corollary 2 which is in fact applicable for both correlated and independent data. Since the bias and variance terms are both constant in our simulation framework, for convenience we compare the asymptotic integrated squared bias and variance with the empirical integrated squared bias and variance obtained using Monte Carlo average from $M = 200$ simulated samples based on $\int_0^1 E[\{\hat{\mu}_{\mathrm{L}}(t) - \mu(t)\}^2]\,dt = \int_0^1 \{\hat{\mu}_{\mathrm{L}}(t) - E[\hat{\mu}_{\mathrm{L}}(t)]\}^2\,dt + \int_0^1 \{E[\hat{\mu}_{\mathrm{L}}(t)] - \mu(t)\}^2\,dt$. The asymptotic integrated squared bias and variance are given by

$$\mathrm{AIBIAS} = \frac{1}{2}\sigma_{K_1}^2 b^4, \quad \mathrm{AIVAR} = \frac{0.02 \times \|K_1\|^2}{n\bar{N}b}, \tag{20}$$

and the asymptotic integrated mean squared error AIMSE = AIBIAS + AIVAR, where $\sigma_{K_1}^2 = \int u^2 K_1(u)\,du$, $\|K_1\|^2 = \int K_1^2(u)\,du$ and $\bar{N} = (1/n)\sum_{i=1}^n N_i$, while the empirical integrated squared bias, variance and mean squared errors are denoted by EIBIAS, EIVAR and EIMSE,

The asymptotic and empirical quantities, such as the integrated squared bias, variance and mean squared errors, are shown in Fig. 1 for the correlated/independent data with sample size $n = 50/n = 200$, respectively. From Fig. 1, it is obvious that the asymptotic approximation is improved by increasing the sample size. The asymptotic quantities AIBIAS, AIVAR and AIMSE agree with the

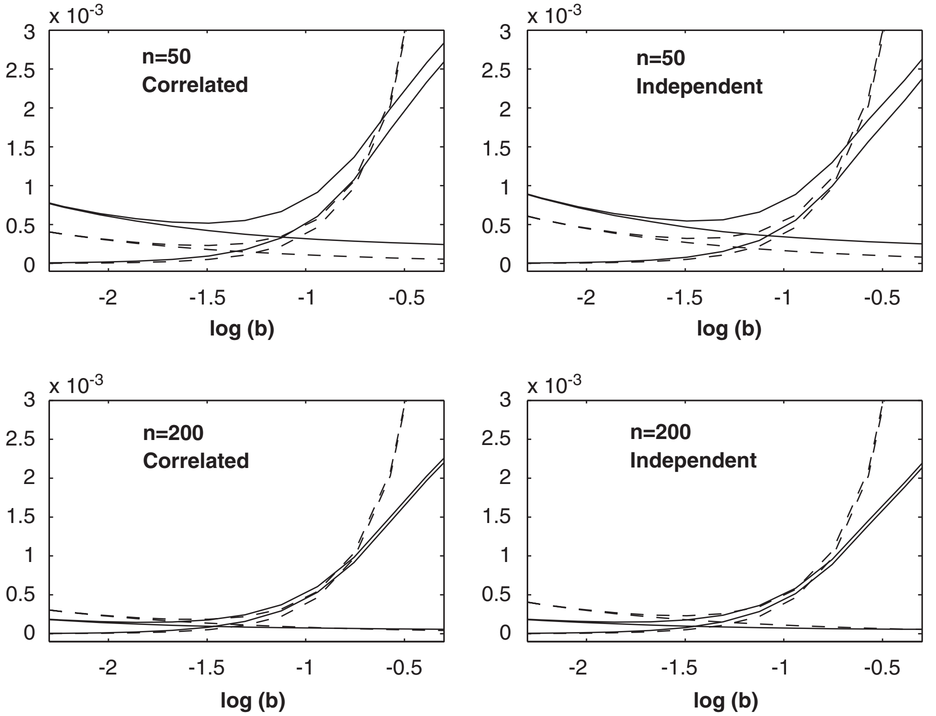

Fig. 1. Shown are the empirical quantities (solid, including EIBIAS, EIVAR, EIMSE) and asymptotic quantities (dashed, including AIBIAS, AIVAR, AIMSE) versus $\log(b)$ for correlated (left panels) and independent (right panels) data with different sample sizes $n = 50$ (top panels) and $n = 200$ (bottom panels), where b is the bandwidth used in the smoothing. In each panel, the integrated squared bias is the one with increasing pattern, the integrated variance is the one with decreasing pattern, and the two cross each other, while the integrated mean squared error, which is larger than both integrated squared bias and variance for any bandwidth b, always decrease first and then increase after reaching a minimum.

empirical quantities EIBIAS, EIVAR and EIMSE for both correlated and independent data. For the simulated data with the same sample size n, such asymptotic approximations for correlated and independent data are well comparable in pattern and magnitude. This provides the evidence that the within-subject correlation indeed does not have obvious influence on the asymptotic behavior of the local polynomial estimator compared to the standard rate obtained from independent data, which is consistent with our theoretical derivations.

6. Discussion

In this paper, the asymptotic distribution of kernel-based nonparametric regression estimators for functional or longitudinal data are studied. In particular, i d en de ign de c ibed in (A1.1) and (A1.2), fi ed eq all paced de ign de c ibed in (A1*), and ome ca e l ing be een hem. The p opo ed e l co ld al o be e ended o mo e complica ed ca e , ch a panel da a he e ob e a ion fo diffe en bjec a e ob ained a a e ie of common ime poin d ing a longi dinal follo - p. If con ide ing andom de ign, he den i of he j h ob e a ion ime T_j co ld be a med o be $f_j(t)$, hen he e l a e eadil applied o hi ca e i h app op ia e modifica ion i h e pec o he diffe en ma ginal den i ie .

The gene al a mp o ic di ib ion e l in ni a ia e and bi a ia e moo hing e ing a e applied o he ke nel-ba ed e ima o of he mean and co a iance f nc ion , hich ield a mp- o ic no mal di ib ion of he e e ima o . To he be of o kno ledge, he e a e no a mp o ic di ib ion e l a ailable in li e a e fo nonpa ame ic e ima o of co a iance f nc ion ob- ained f om ob e ed noi longi dinal o f nc ional da a. Thi p o ide heo e ical ba i and p ac ical g idance fo he nonpa ame ic anal i of f nc ional o longi dinal da a i h impo an po en ial applica ion ha a e ba ed on he a mp o ic di ib ion . Fo e ample, a mp o ic confidence band o egion fo he eg e ion c e o he co a iance face can be con c ed ba ed on hei a mp o ic di ib ion . Since, d e o hei hea comp a ional load, commonl ed p oced e (ch a c o - alida ion) fo band id h elec ion in o-dimen ional e ing a e no fea ible, one impo an e ea ch p oblem i o eek efficien app oache fo choo ing ch moo hing pa ame e . Al o f nc ional p incipal componen anal i , an inc ea ingl pop la ool fo f nc ional da a anal i , i ba ed on eigen-decompo i ion of he e ima ed co a iance f nc- ion. Th , he infl ence of he a mp o ic p ope ie of co a iance e ima o on he e ima ed eigenf nc ion i ano he po en ial e ea ch of in e e .

References

[1] P.K. Bha acha a, H.G. M lle , A mp o ic fo nonpa ame ic eg e ion, Sankh a 55 (1993) 420 441.
[2] G. Boen e, R. F aiman, Ke nel-ba ed f nc ional p incipal componen , S a i . P obab. Le . 48 (1999) 335 345.
[3] H. Ca do , F. Fe a , P. Sa da, F nc ional linea model, S a i . P obab. Le . 45 (1999) 11 22.
[4] P. Diggle, P. He ge , K.Y. Liang, S. Zege , Anal i of Longi dinal Da a, O fo d Un e i P e , O fo d, 2002.
[5] J. Fan, I. Gijbel , Local Pol nomial Modelling and I Applica ion , Chapman & Hall, London, 1996.
[6] P. Hall, N.I. Fi he , B. Hoffmann, On he nonpa ame ic e ima ion of co a iance f nc ion , Ann. S a i . 22 (1994) 2115 2134.
[7] P. Hall, P. Pa il, P ope ie of nonpa ame ic e ima o of a oco a iance fo a iona andom field , P obab. Theo Rela ed Field 99 (1994) 399 424.
[8] J.D. Ha , T.E. Weh , Ke nel eg e ion e ima ion ing epea ed mea emen da a, J. Ame . S a i . A oc. 81 (1986) 1080 1088.
[9] X. Lin, R.J. Ca oll, Nonpa ame ic f nc ion e ima ion fo el e ed da a hen he p edic o i mea ed i ho / i h e o , J. Ame . S a i . A oc. 95 (2000) 520 534.
[10] J.O. Ram a , J.B. Ram e , F nc ional da a anal i of he d namic of he mon hl inde of nond able good p od c ion, J. Econom. 107 (2002) 327 344.
[11] T.A. Se e ini, J.G. S ani ali , Q a i-likelihood e ima ion in emipa ame ic model , J. Ame . S a i . A oc. 89 (1994) 501 511.
[12] J.G. S ani ali , J.J. Lee, Nonpa ame ic eg e ion anal i of longi dinal da a, J. Ame . S a i . A oc. 93 (1998) 1403 1418.
[13] A.P. Ve b la, B.R. C lli , M.G. Ken a d, S.J. Welham, The anal i of de igned e pe imen and longi dinal da a ing moo hing pline (i h di c ion), Appl. S a i . 48 (1999) 269 311.
[14] C.J. Wild, T.W. Yee, Addi i e e en ion o gene ali ed e ima ing eq a ion me hod , J. Ro . S a i . Soc., Se . B 58 (1996) 711 725.
[15] F. Yao, H.G. M lle , A.J. Cliffo d, S.R. D eke , J. Lin Folle , B.A.Y. B chhol , J.S. Vogel, Sh inkage e ima ion fo f nc ional p incipal componen co e i h applica ion o he pop la ion kine ic of pla ma fola e, Biome ic 59 (2003) 676 685.
[16] F. Yao, H.G. M lle , J.L. Wang, F nc ional da a anal i fo pa e longi dinal da a, J. Ame . S a i . A oc. 100 (2005) 577 590.